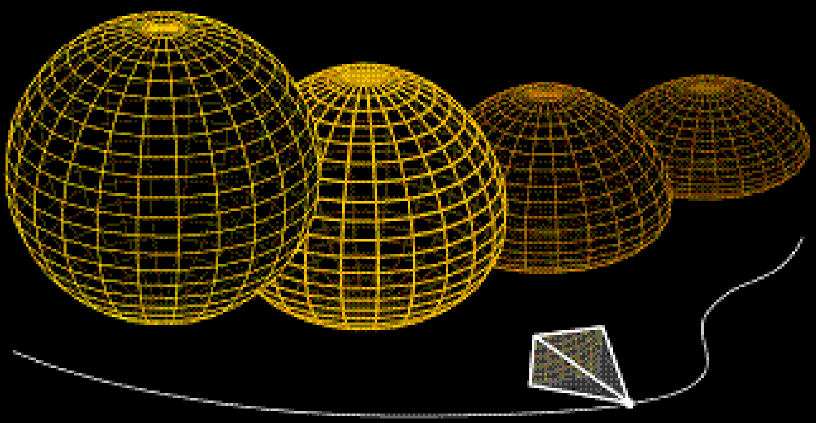
Large-scale optimisation of bilinear models in Computer Vision (and more)



Alessio Del Bue

PostDoc Senior

PLUS LAB

Istituto Italiano di Tecnologia

Genova, 8th of June, 2010

The PLUS crew

- **Vittorio Murino** (*Director*)
- Manuele Bicego (*Team Leader*)
- Marco Cristani (*Team Leader*)
- Alessio Del Bue (*PostDoc Senior*)
- Reza Sabzevari (*Ph.D. student*)
- Pietro Salvagnini (*Ph.D. student*)
- Michele Stoppa (*Ph.D. Student*)

Lab

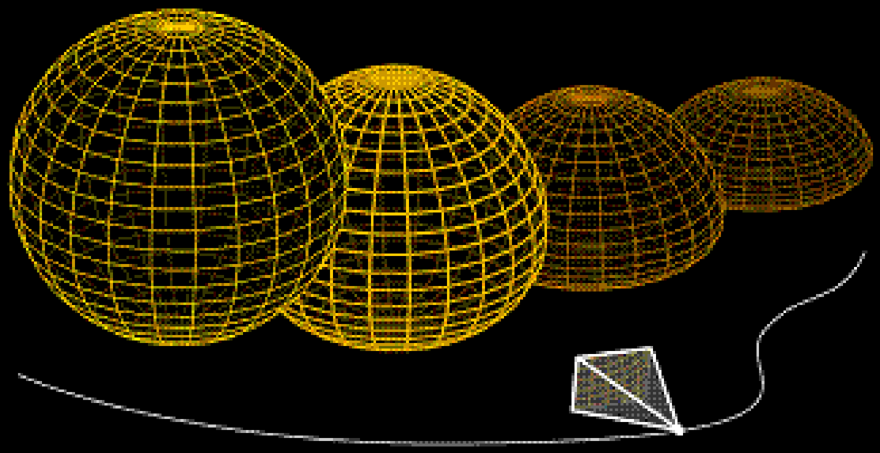
- Social Signal Processing
  - Attention, emotions, expressions
  - Behaviors (audio, video)
  - Dialogs
  - Events (audio, video)
  - Gaze/pose estimation
- Video analytics
  - Tracking, tagging (re-ID: individuals, groups, crowd)
  - Subjective surveillance
  - Detection/classification
  - Retrieval/HCI, visualization, video/data mining

PLUS Lab

- Biometrics
  - Face recognition: non cooperative, at distance
  - Super resolution
  - Gait
  - Tagging

- Sensors/Data Fusion
  - Sensors Network
  - Embedded computer vision, smart sensors
  - Multi-modal sensors: stereo, 3D, omni-directional, IR, RFID, proximity, etc.

Computer Vision in a Non-Rigid World



In collaboration with:

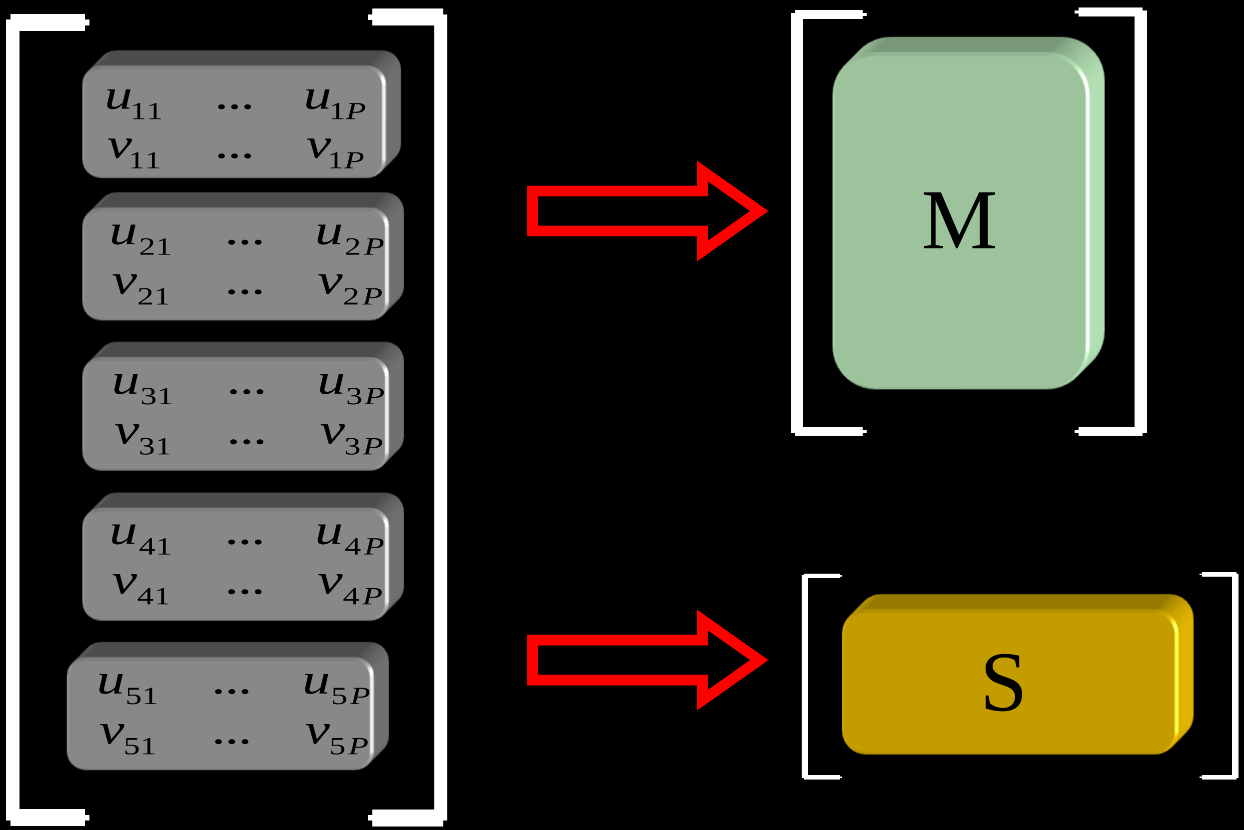
Lourdes Agapito - QMUL/London
Marco Paladini - QMUL/London
Xavier Llado – UdG/Girona
Joao Xavier - ISR-IST/Lisbon
Joao Fayad - QMUL/London
Marko Stosic - ISR-IST/Lisbon
Fabrizio Smeraldi - QMUL/London

Non-rigid structure from motion

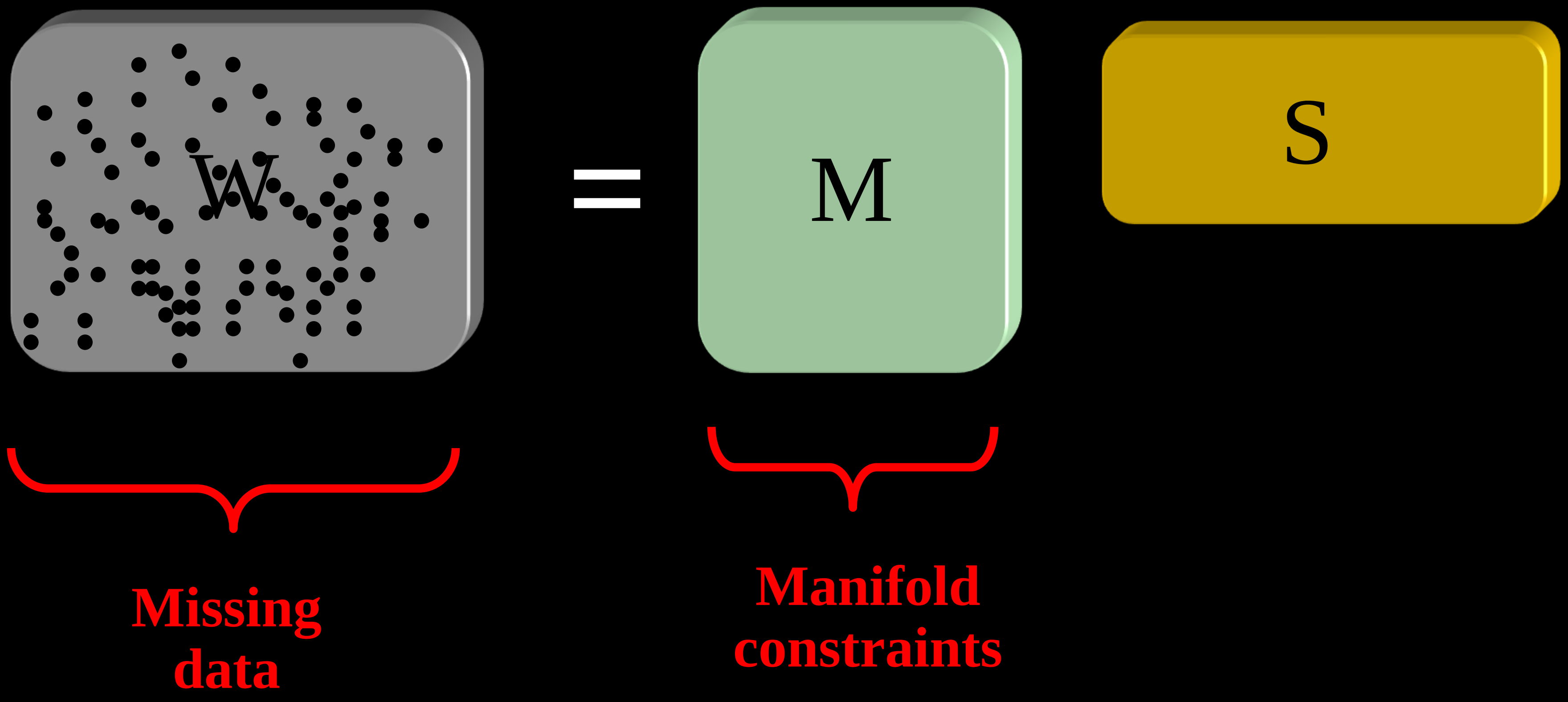
To extract non-rigid 3D models from uncalibrated video sequences

Global 3D Reconstruction

Aim: To design **model-free** algorithms which exploit the **complete** image data of the object shape.



A computational framework



To infer from the data both the components of bilinear model with missing data and non-linear constraints

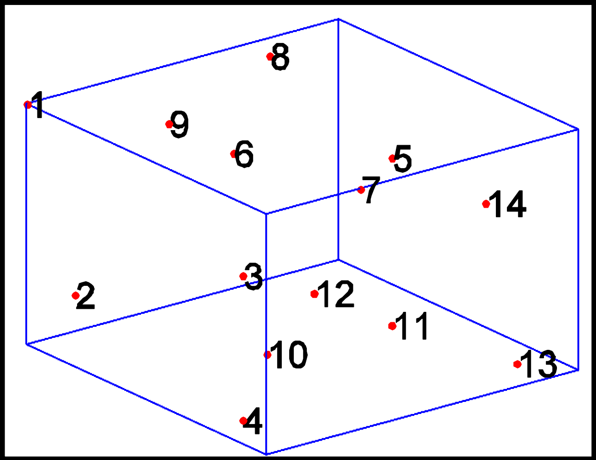
Missing Data and Manifold Constraints - "The BALM"

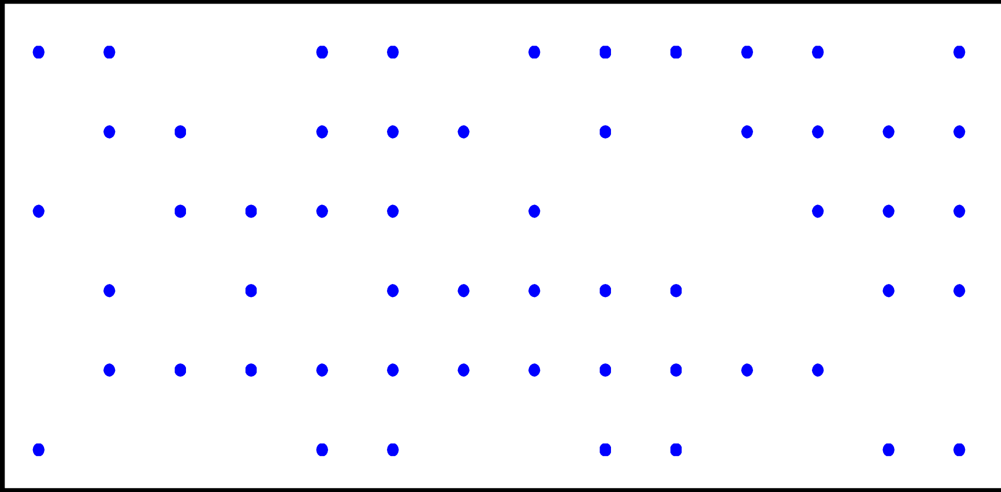
A. Del Bue, J. Xavier, L. Agapito, and M. Paladini, "Bilinear factorization via Augmented Lagrange Multipliers," ECCV 2010 Poster.

1^{st} Reason
Why using manifold constraints?

The projection to a motion manifold was first introduced in the rigid case and in the presence of missing data to obtain a robust 3D reconstruction.

Example:

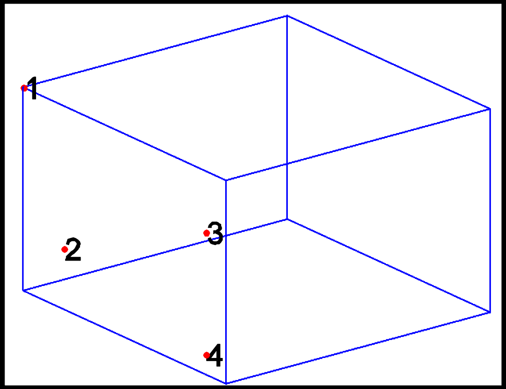


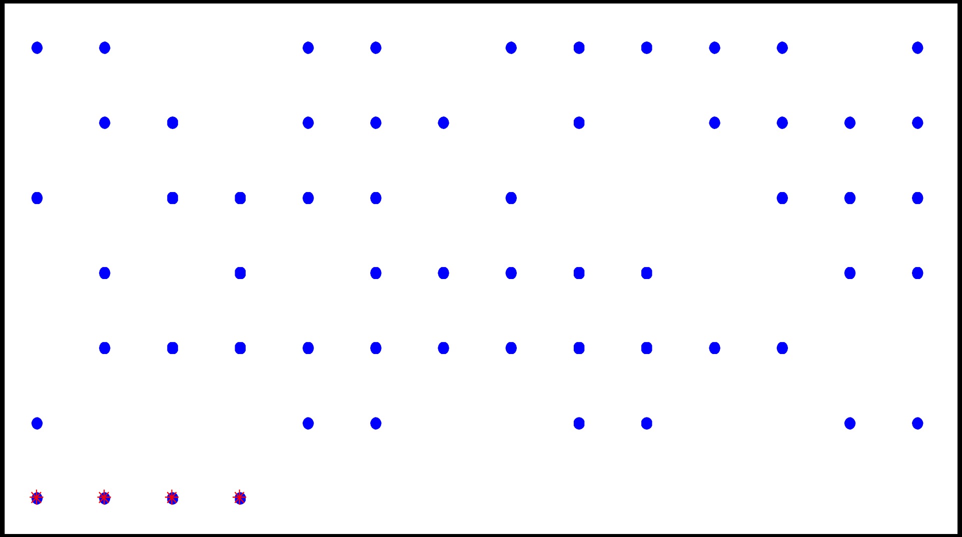


Marques and Costeira. Estimating 3D shape from degenerate sequences with missing data. CVIU 2009

Why do a Motion Manifold really help? Degenerate sequences

Example:



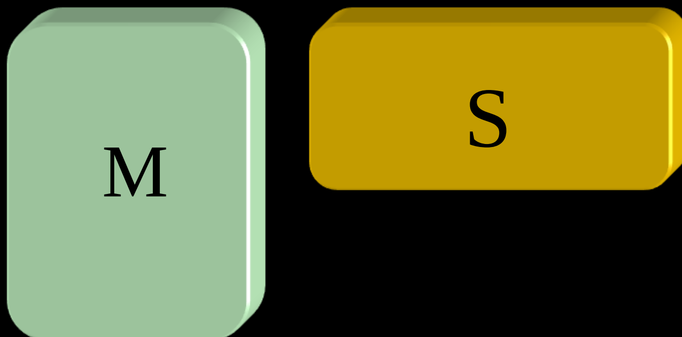


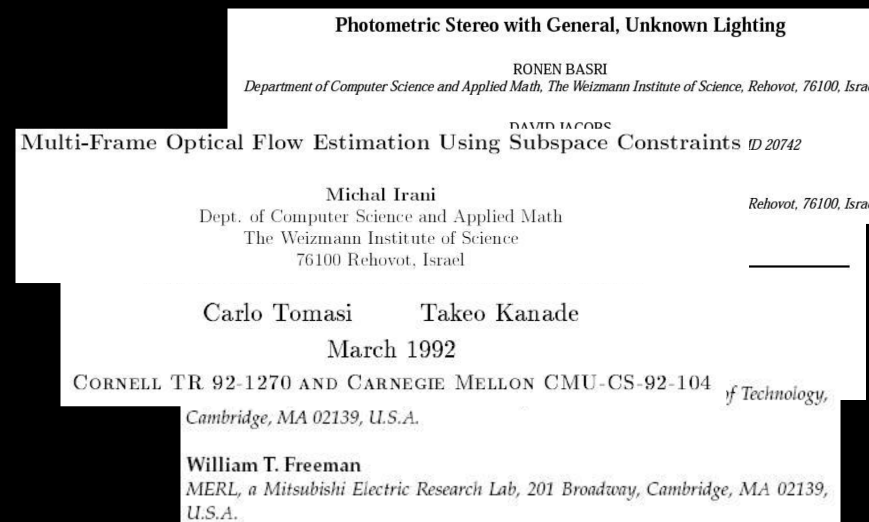
If a single frame contains a configuration of 2D points that is degenerate, the accuracy of standard SfM algorithms decrease dramatically.

2nd Reason

Several problems, 1 bilinear solution

W =




Photometric Stereo with General, Unknown Lighting

RONEN BASRI
Department of Computer Science and Applied Math, The Weizmann Institute of Science, Rehovot, 76100, Israel

Multi-Frame Optical Flow Estimation Using Subspace Constraints

Michal Irani
Dept. of Computer Science and Applied Math
The Weizmann Institute of Science
76100 Rehovot, Israel

Rehovot, 76100, Israel

Carlo Tomasi Takeo Kanade
March 1992
Cornell TR 92-1270 and Carnegie Mellon CMU-CS-92-104

Cambridge, MA 02139, U.S.A.

William T. Freeman
MERL, a Mitsubishi Electric Research Lab, 201 Broadway, Cambridge, MA 02139, U.S.A.

Measurements are function of bilinear components

- Structure from motion (Rigid, deformable, articulated)
- Photometry (M: light field direction – S: shape normals and albedo)
- Style and Content (i.e. M: calligraphy – S: letter)
- Structure from Sound (M: sound direction – S: mic location)
- Optical flow (M: 2D displacement – S: image gradients)

One optimisation problem

$$\begin{aligned}&\text{minimize} \quad \|W(z) - MS\|^2\\&\text{subject to } M_i \in \mathcal{M}, \quad i = 1, \ldots, f,\end{aligned}$$

Each columns of M lies on a specific *manifold* given by the problem.

$$(Y(z))_{ij} := \begin{cases} Y_{ij}, & \text{if } (i,j) \in \mathcal{O} \\ Z_{ij}, & \text{if } (i,j) \notin \mathcal{O} \end{cases}$$

With missing data, we have also to estimate the missing entries Z_{ij}

The novel concept

Why designing different algorithms if the task is the same?

"The main difference between different factorization problems is the manifold on which the solution lies. Thus, intuitively, it should be possible to construct an unified optimization framework in which a change of the manifold constraint just implies replacing an inner module of the algorithm"

We introduce a reformulation that decouples the core bilinear aspect of the problem from the manifold specificity.

$$\begin{aligned} &\text{minimize} \quad \|W(z) - MS\|^2 \\ &\text{subject to } M_i \in \mathcal{M}, \quad i = 1, \ldots, f, \end{aligned}$$

Algorithm structure

Algorithm 1 Bilinear factorization via Augmented Lagrange Multipliers (BALM)

set $k = 0$ and $\epsilon_{\text{best}} = +\infty$
initialize $\sigma^{(0)}$, $R^{(0)}$, $\gamma > 1$ and $0 < \eta < 1$
initialize $z^{(0)}$, $S^{(0)}$ and $M^{(0)}$
repeat
  solve

$$\left(z^{(k+1)}, S^{(k+1)}, M^{(k+1)}, N^{(k+1)}\right) = \begin{array}{ll} \operatorname{argmin} & L_{\sigma^{(k)}}(z, S, M, N; R^{(k)}) \\ \text{subject to } N_i \in \mathcal{M}, & i = 1, \dots, f, \end{array}$$

  using the iterative Gauss-Seidel scheme described in **Algorithm 2**
  compute $\epsilon = \left\| M^{(k+1)} - N^{(k+1)} \right\|^2$
  if $\epsilon < \eta\, \epsilon_{\text{best}}$
    $R^{(k+1)} = R^{(k)} - \sigma^{(k)} \left(M^{(k+1)} - N^{(k+1)} \right)$
    $\sigma^{(k+1)} = \sigma^{(k)}$
    $\epsilon_{\text{best}} = \epsilon$
  else
    $R^{(k+1)} = R^{(k)}$
    $\sigma^{(k+1)} = \gamma \sigma^{(k)}$
  endif
  update $k \leftarrow k + 1$
until some stopping criterion

Which can be summarised in three steps…

1. Replicate variables

$$\begin{aligned} \text{minimize} \quad & \|W(z) - MS\|^2 \\ \text{subject to } & M_i = N_i, \quad i = 1, \ldots, f \\ & N_i \in \mathcal{M}, \quad i = 1, \ldots, f. \end{aligned}$$

The replicated set of variables N separates the manifold restriction from the core bilinear problem.

$$L_\sigma(z, S, M, N; R) = \|W(z) - MS\|^2 - \sum_{i=1}^{f} \operatorname{tr}\left(R_i^\top (M_i - N_i)\right) + \frac{\sigma}{2} \sum_{i=1}^{f} \|M_i - N_i\|^2 .$$

The cost function optimised in based on the ALM method.

2. Solve separate problems

$$N^{[l+1]} = \operatorname{argmin} \quad L_{\sigma(k)}\left(z^{[l]}, S^{[l]}, M^{[l]}, N; R^{(k)}\right)$$
$$\text{subject to } N_i \in \mathcal{M}, \quad i = 1, \ldots, f,$$

2a. Solve for the *manifold* projection given *M* (custom for each problem)

$$\left(S^{[l+1]}, M^{[l+1]}\right) = \operatorname{argmin} L_{\sigma(k)}\left(z^{[l]}, S, M, N^{[l+1]}; R^{(k)}\right)$$

2b. Solve for the bilinear factors given *N* using Gauss-Siedel

3. Missing data imputation

$$z^{[l+1]} = \operatorname{argmin} L_{\sigma(k)} \left(z, S^{[l+1]}, M^{[l+1]}, N^{[l+1]}; R^{(k)} \right)$$

Given *M* and *S* fill the missing entries in W as $Z_{ij} = M_i S_j$

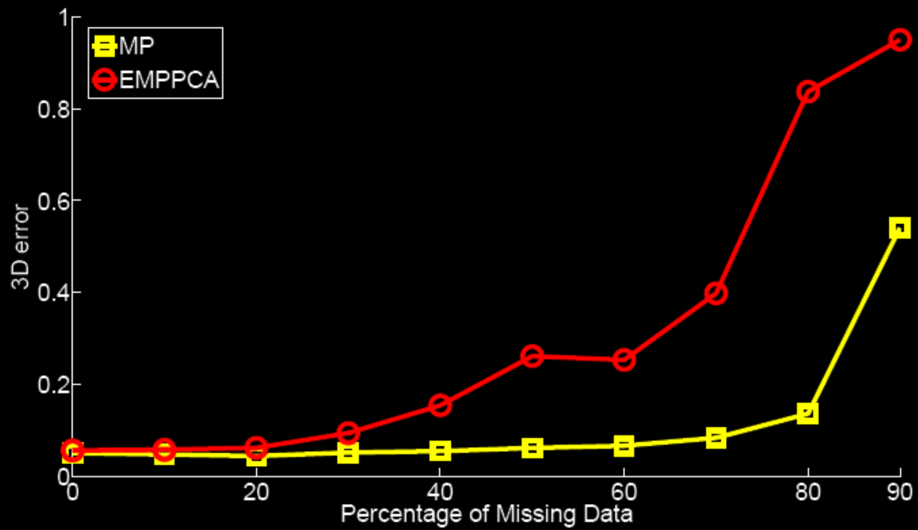
The algorithm iterates until a given stop criteria (update on the parameters). Under some restrictive conditions Augmented Lagrangian methods are locally convergent to a minima.

Experiments on NRSfM

Results – Missing Data

Soon on: http://www.isr.ist.utl.pt/~adb/

More experiments



Synthetic test against the state of the art algorithm shows that the metric projections are very robust with high percentages of missing data.

To notice that the state of the art method (EM-PPCA) uses priors over the distributions of the deformation weights and linear dynamics model of the motion.

The metric projection algorithm uses exclusively the geometric constraints of the problem.

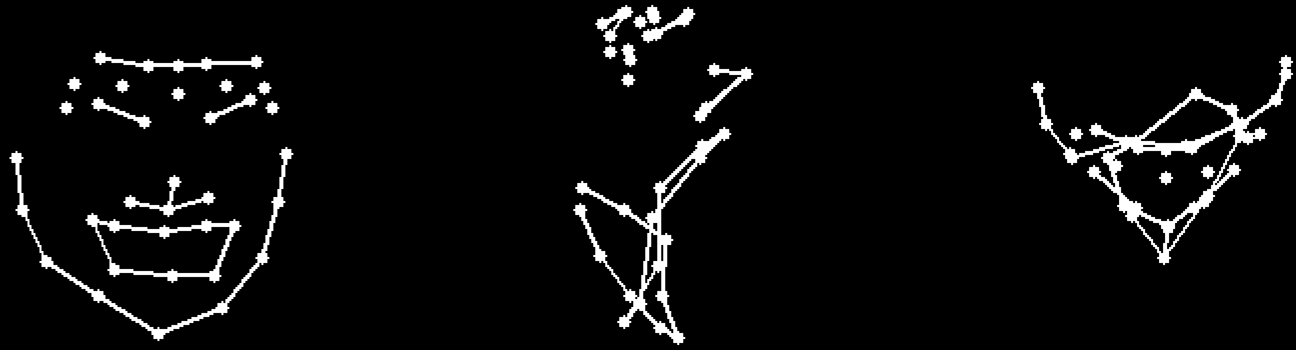
Structure from Motion with Priors

A. Del Bue, "Adaptive Metric Registration of 3D Models to Non-rigid Image Trajectories“ ECCV 2010 Oral

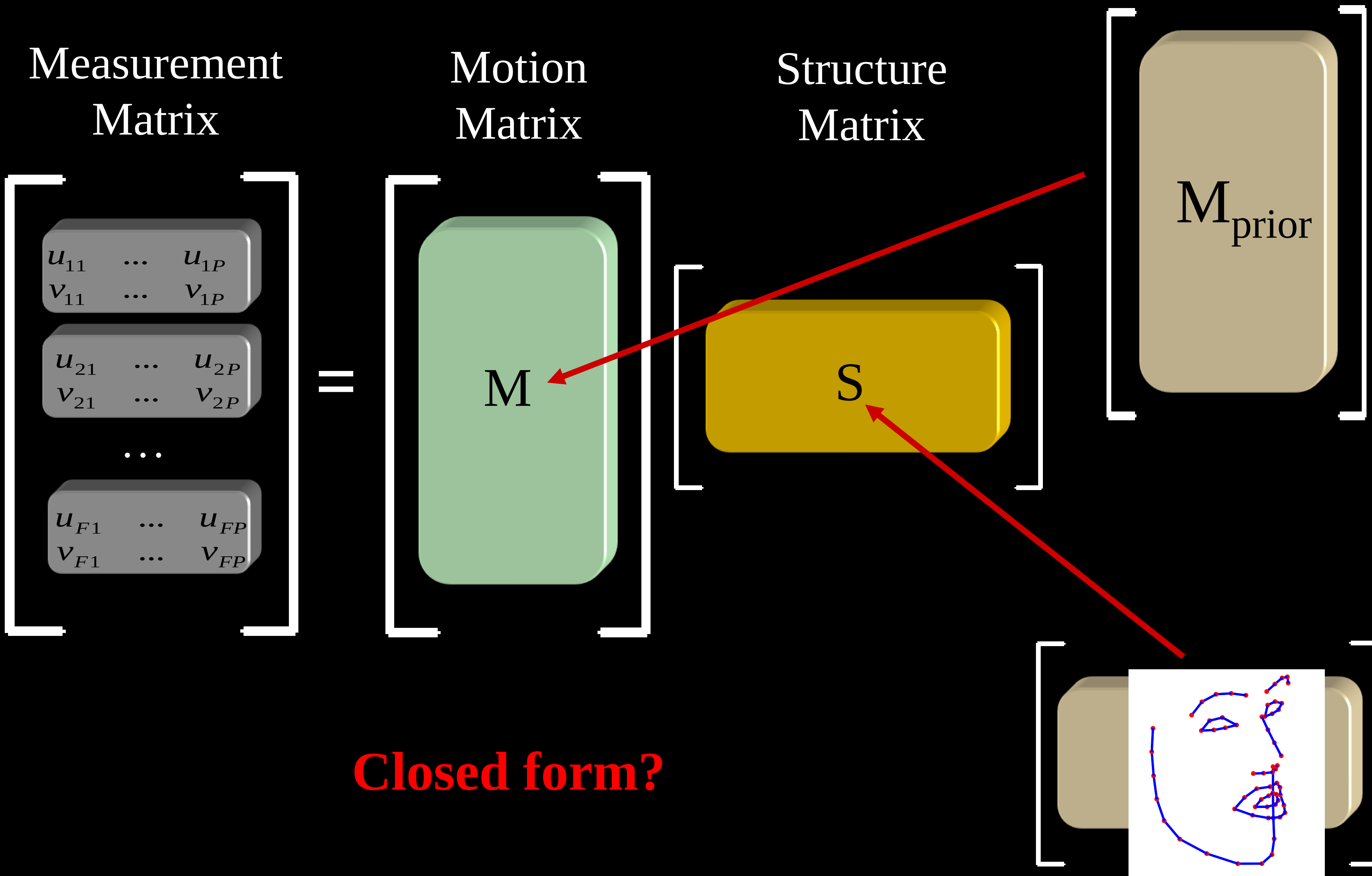
Degenerate Motion

More often then expected, video sequences do not provide sufficient rigid motion in order to reconstruct correct 3D shapes.

Non-rigid SfM needs strong rigid motion to avoid ambiguities



Is it possible to include priors?



GSVD!

$$\begin{bmatrix} \begin{matrix} u_{11} & \cdots & u_{1P} \\ v_{11} & \cdots & v_{1P} \end{matrix} \\ \begin{matrix} u_{21} & \cdots & u_{2P} \\ v_{21} & \cdots & v_{2P} \end{matrix} \\ \cdots \\ \begin{matrix} u_{F1} & \cdots & u_{FP} \\ v_{F1} & \cdots & v_{FP} \end{matrix} \end{bmatrix} \Rightarrow U \quad D_U \quad X^T$$

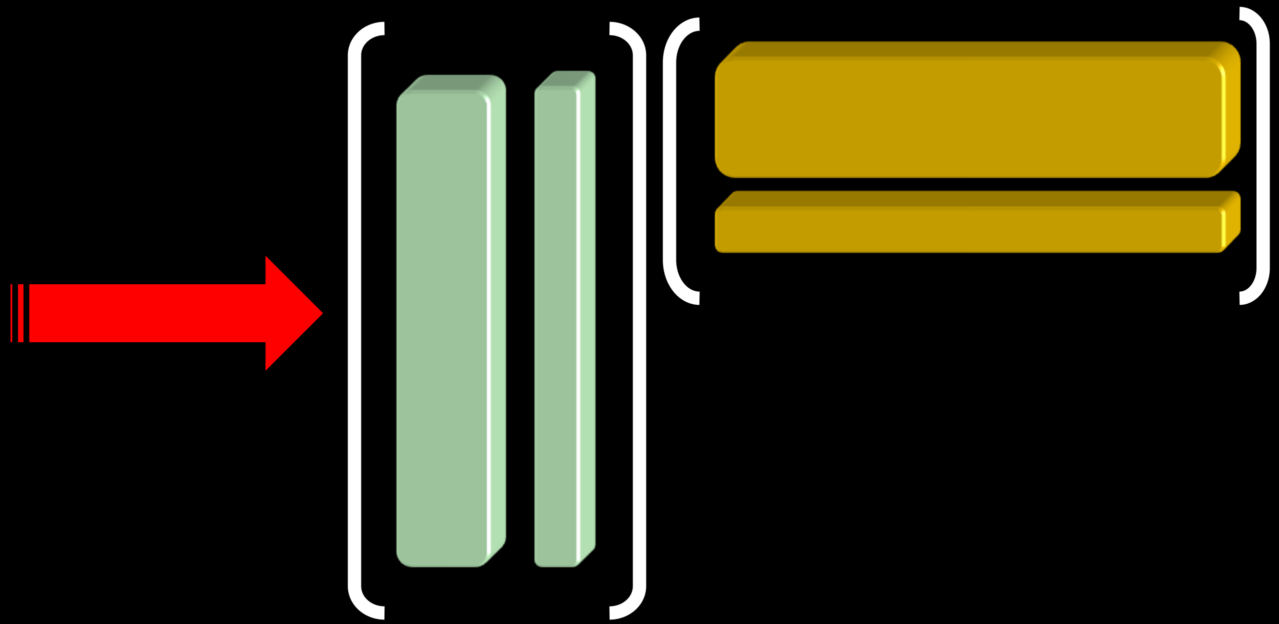
Rank Constrained!

$$\begin{bmatrix} S_{prior} \end{bmatrix} \Rightarrow V \quad D_V \quad X^T$$

Closed form affine solution

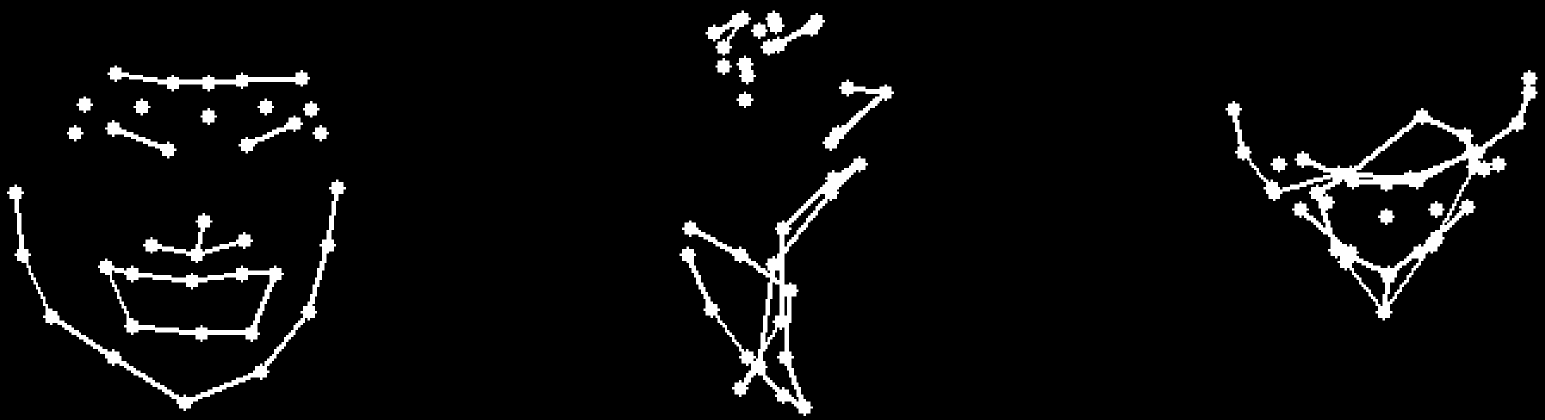
$$W = U\, D_U\, X^T$$

$$S_P = V\, D_V\, X^T$$

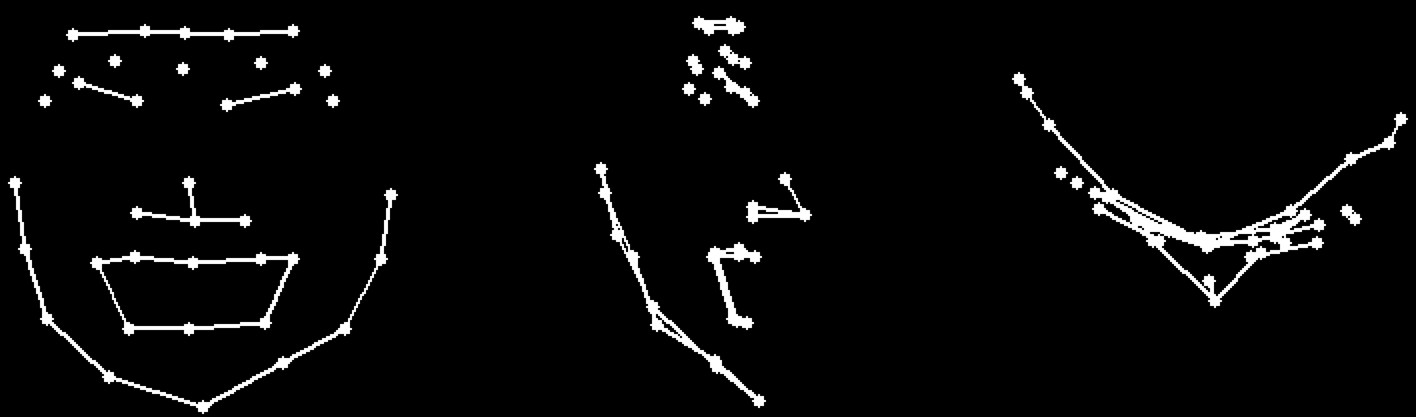


Metric solution is then achieved with a reformulation of rigid factorisation, again the solution can be computed in closed form.

No Prior



With Prior



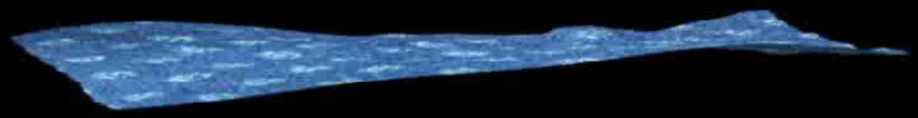
3D Face Reconstruction

Check: www.isr.ist.utl.pt/~adb/research-projects/structure-from-motion-with-shape-priors/

Highly Non-linear deformations: A piecewise approach

J. Fayad, L. Agapito, and A. Del Bue, "Piecewise Quadratic Reconstruction of Non-Rigid Surfaces from Monocular Sequences", ECCV 2010 Poster

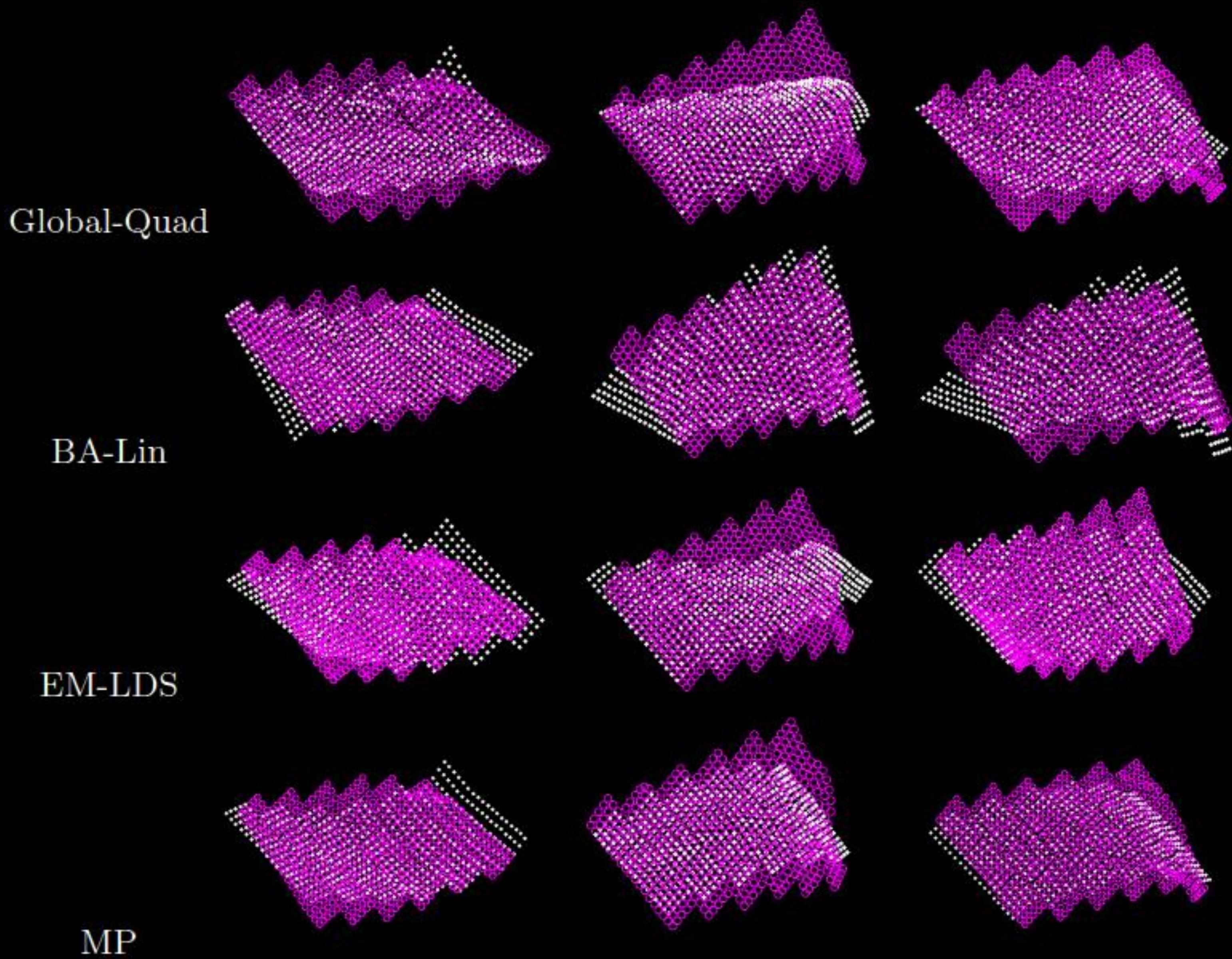
What about strongly deforming bodies?



Standard SfM approaches fail when dealing with strong deformations such as paper bending or torsion.

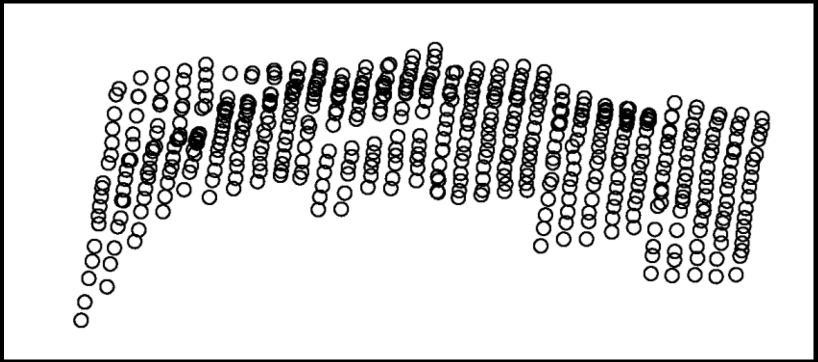
However, a globally deforming body can be reasonably approximated by locally linear/non-linear models.

Results with previous NRSfM methods

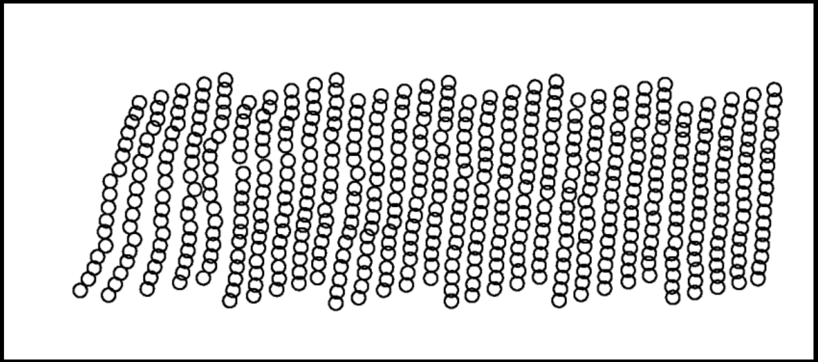


A pipeline for the reconstruction - 1

1. Extract a "mean" shape using standard rigid SfM from few frames of the sequence.

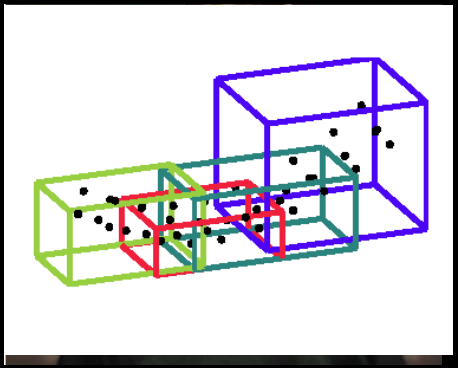


2. Compute an (isometric) low-dimensional embedding of the surface .



A pipeline for the reconstruction - 2

3. Form the cluster using nearest neighbours in the "flattened" space with consistent overlay.

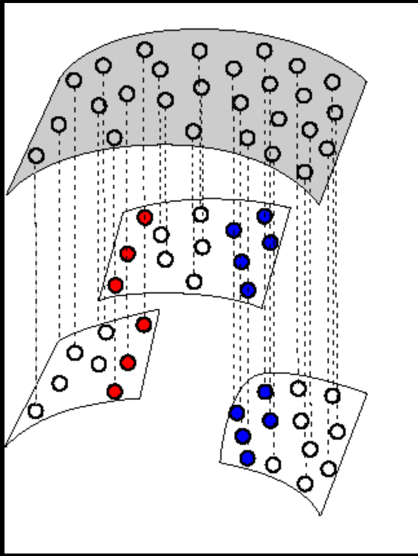


4. Optimise each patch with a quadratic or linear deformation model minimizing the reprojectiong error.

$$\min_{\mathbf{q}_i,\mathbf{t}_i,\mathtt{L}_i,\mathtt{Q}_i,\mathtt{C}_i} \sum_{i,j}^{f,p} ||\mathbf{w}_{ij} - \mathtt{R}_i(\mathbf{q}_i)\,[\mathtt{L}_i\,\mathtt{Q}_i\,\mathtt{C}_i]\,\mathbf{S}_j - \mathbf{t}_i||^2$$

A pipeline for the reconstruction - 3

5. Connect and align patches given the overlapping points in each cluster. Solve for ambiguities.



6. Perform a global optimization to reduce the discrepancy between patches.

$$\sum_{i,j}^{f,p} \sum_{n \in \Theta_j} \left|\left| \mathbf{w}_{ij}^{(n)} - \hat{\mathbf{w}}_{ij}^{(n)} \right|\right|^2 + \\ + \ \eta \sum_{k \in \Theta_j / \{n\}} \left|\left| \hat{\mathbf{X}}_{ij}^{(n)} - \hat{\mathbf{X}}_{ij}^{(k)} \right|\right|^2$$

Flag Experiment

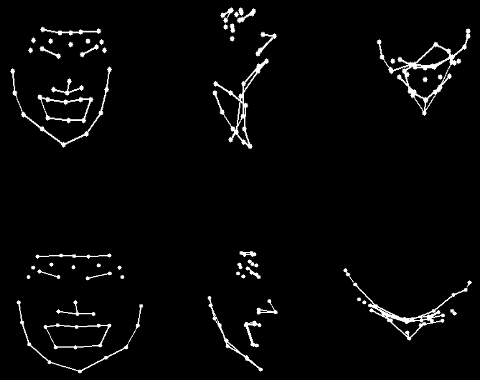
Soon on: http://www.isr.ist.utl.pt/~adb/

Bending Paper Experiment

Soon on: http://www.isr.ist.utl.pt/~adb/

To summarise

Bilinear reconstruction with manifold constraints and high percentage of missing data (AKA matrix completion).



Structure from Motion with priors.

Piecewise reconstruction of strongly deformable surfaces.

Thank you!

For more info and results: http://www.isr.ist.utl.pt/~adb/